WHAT IS CLAIMED IS:

1. A method of extracting data from a file, the method comprising:

receiving a request to extract one or more data records from the file;

identifying the data records within the file, without using prior knowledge of the structure of the file; and

extracting the data records.

2. The method of claim 1, wherein the data records contain at least one tag and text, and wherein the method further comprises the step of separating each tag from the text by adding a space between the tag and the text.

3. The method of claim 2, further comprising converting the text into a list that contains one or more words, wherein each word is associated with its position in the list.

4. The method of claim 3, further comprising:

evaluating each word based on a meaning of the word and based on the position of the word in the list; and

based on the evaluation, calculating a score for the word.

5. The method of claim 3, further comprising:

converting the text into a list that contains one or more phrases;

determining whether one of the words, at a specified position in the word list, matches one of the phrases in the phrase list; and

based on that determination, calculating a score for the phrase.

6. The method of claim 3, further comprising:

determining whether one of the words, at a specified position in the word list, matches a predefined expression; and

based on that determination, calculating a score for the predefined expression.

7. The method of claim 6, wherein the predefined expression is a dollar value.

8. The method of claim 4, further comprising:

identifying a region of interest within the word list by determining which region has a highest grade; and

using the identified region of interest to create a set of sub-regions of the word list.

9. The method of claim 8, further comprising partitioning one of the sub- regions into one or more record regions, wherein each record region contains text that corresponds to a single record.

10. The method of claim 9, further comprising extracting one or more attribute values from the record region.

11. The method of claim 10, wherein extracting the attribute values comprises extracting one or more phrases.

12. The method of claim 10, wherein extracting the attribute value comprises extracting a dollar amount

13. The method of claim 10, further comprising writing a set of records to a file.

14. The method of claim 1, further comprising writing the extracted data record to an output file.

15. An apparatus for extracting data from a file, comprising:

a computer; and

a computer program, performed by the computer, for receiving a request to extract one or more data records from the file, identifying the data records within the file, without using prior

knowledge of the structure of the file, and extracting the data records.

16. The apparatus of claim 15, wherein the data records contain at least one tag and text, and wherein the apparatus further comprises a computer program, performed by the computer, for separating each tag from the text by adding a space between the tag and the text.

17. The apparatus of claim 16, further comprising a computer program, performed by the computer, for converting the text into a list that contains one or more words, wherein each word is associated with its position in the list.

18. The apparatus of claim 17, further comprising:

a computer program, performed by the computer, for evaluating each word based on a meaning of the word and based on the position of the word in the list, and based on the evaluation, calculating a score for the word.

19. The apparatus of claim 17, further comprising:

a computer program, performed by the computer, for converting the text into a list that contains one or more phrases, determining whether one of the words, at a specified position in the word list, matches one of the phrases in the phrase list, and based on that determination, calculating a score for the phrase.

20. The apparatus of claim 17, further comprising:

a computer program, performed by the computer, for determining whether one of the words, at a specified position in the word list, matches a predefined expression, and based on that determination, calculating a score for the predefined expression.

21. The apparatus of claim 20, wherein the predefined expression is a dollar value.

22. The apparatus of claim 18, further comprising:

a computer program, performed by the computer, for identifying a region of interest

within the word list by determining which region has a highest grade, and using the identified region of interest to create a set of sub-regions of the word list.

23. The apparatus of claim 22, further comprising a computer program, performed by the computer, for partitioning one of the sub- regions into one or more record regions, wherein each record region contains text that corresponds to a single record.

24. The apparatus of claim 23, further comprising a computer program, performed by the computer, for extracting one or more attribute values from the record region.

25. The apparatus of claim 24, wherein extracting the attribute values comprises extracting one or more phrases.

26. The apparatus of claim 24, wherein extracting the attribute value comprises extracting a dollar amount

27. The apparatus of claim 26, further comprising a computer program, performed by the computer, for writing a set of records to a file.

28. The apparatus of claim 15, further comprising a computer program, performed by the computer, for writing the extracted data record to an output file.

29. An article of manufacture comprising a computer program carrier readable by a computer and embodying one or more instructions executable by the computer to perform the method of extracting data from a file, the method comprising:

receiving a request to extract one or more data records from the file;

identifying the data records within the file, without using prior knowledge of the structure of the file; and

extracting the data records.

30. The article of manufacture of claim 29, wherein the data records contain at least one tag and text, and wherein the article of manufacture further comprises the step of separating each tag from the text by adding a space between the tag and the text.

31. The article of manufacture of claim 30, further comprising converting the text into a list that contains one or more words, wherein each word is associated with its position in the list.

32. The article of manufacture of claim 31, further comprising:

evaluating each word based on a meaning of the word and based on the position of the word in the list; and

based on the evaluation, calculating a score for the word.

33. The article of manufacture of claim 31, further comprising:

converting the text into a list that contains one or more phrases;

determining whether one of the words, at a specified position in the word list, matches one of the phrases in the phrase list; and

based on that determination, calculating a score for the phrase.

34. The article of manufacture of claim 31, further comprising:

determining whether one of the words, at a specified position in the word list, matches a predefined expression; and

based on that determination, calculating a score for the predefined expression.

35. The article of manufacture of claim 34, wherein the predefined expression is a dollar value.

36. The article of manufacture of claim 32, further comprising:

identifying a region of interest within the word list by determining which region has a highest grade; and

using the identified region of interest to create a set of sub-regions of the word list.

37. The article of manufacture of claim 36, further comprising partitioning one of the sub- regions into one or more record regions, wherein each record region contains text that corresponds to a single record.

38. The article of manufacture of claim 37, further comprising extracting one or more attribute values from the record region.

39. The article of manufacture of claim 38, wherein extracting the attribute values comprises extracting one or more phrases.

40. The article of manufacture of claim 38, wherein extracting the attribute value comprises extracting a dollar amount

41. The article of manufacture of claim 38, further comprising writing a set of records to a file.

42. The article of manufacture of claim 29, further comprising writing the extracted data record to an output file.